(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 1 365 417 A1

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
26.11.2003 Bulletin 2003/48

(51) Int Cl.7: G11C 16/10, G11C 11/56

(21) Application number: 02425293.4

(22) Date of filing: 13.05.2002

(84) Designated Contracting States:
AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE TR
Designated Extension States:
AL LT LV MK RO SI

(71) Applicant: STMicroelectronics S.r.l.
20041 Agrate Brianza (Milano) (IT)

(72) Inventors:
- Pascucci, Luigi
  20099 Sesto San Giovanni (Milano) (IT)
- Rolandi, Paolo
  27058 Voghera (Pavia) (IT)
- Riva, Marco
  24040 Lallio (Bergamo) (IT)

(74) Representative: Botti, Mario
Botti & Ferrari S.r.l.,
Via Locatelli, 5
20124 Milano (IT)

(54) Programming method of the memory cells in a multilevel non-volatile memory device

(57) The invention relates to a method for programming a non-volatile memory device of the multi-level type, comprising a plurality of transistor cells grouped into memory words and conventionally provided with gate and drain terminals. The method applies different drain voltage values at different threshold values. Such values are directly proportional to the threshold levels to be attained by the individual memory word bits, and effective to provide for a simultaneous attainment of the levels, in a seeking-to manner, of the levels at the end of a limited number of pulses.

Advantageously, a constant gate voltage value is concurrently applied to the gate terminals of said cells, such that the cell programming time is unrelated to the threshold level sought.

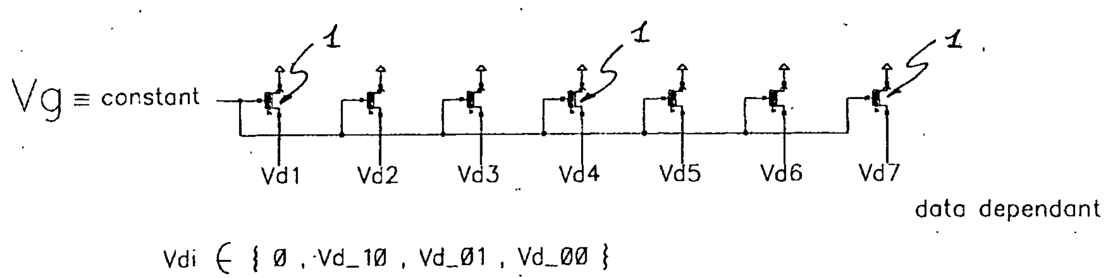


FIG. 2

EP 1 365 417 A1

Description

Field of Application

[0001] The present invention relates to a method for programming the memory cells of a multi-level non-volatile memory device.

[0002] More particularly, the invention relates to a method of programming a non-volatile memory device of the multi-level type, which device comprises a plurality of transistor cells grouped into memory words and provided with conventional gate and drain terminals.

[0003] The invention relates, particularly but not exclusively, to a multi-level non-volatile memory device monolithically integrated in a semiconductor, and the following description is made with reference to this application field for convenience of illustration only.

Prior Art

[0004] As it is well known, in a two-level memory device that employs a floating gate MOS transistor as elementary cell, the threshold voltage of the cell can be modulated to establish either of two logic states. A first logic state (logic "1") corresponds to a situation of the floating gate containing no charge, as is typical of a virgin or an erased cell, for example. Another logic state (logic "0") corresponds to the floating gate storing a sufficient number of electrons to produce a dramatic rise of its threshold, thereby denoting the programmed state of the cell.

[0005] A current-read method can be used for reading a memory cell 1, which consists of applying a read voltage Vread to the gate terminal of the cell and reading the current that flows through the cell:

- if the cell is a written cell, its threshold voltage will be higher than the read voltage Vread, so that no current will flow through the cell; otherwise
- if the cell is an erased cell, its threshold voltage must be adequate to admit a current through the cell.

[0006] To read information contained in memory cells of this type, a sense amplifier is used which compares the cell current with a reference value, thus converting the analog information of the addressed data in the cell, i.e. the value of the cell threshold voltage, to digital information, i.e. to a logic "0" or a logic "1".

[0007] Non-volatile memory devices, particularly those of the EEPROM and FLASH types, are specially adapted to store up large amounts of data, and are widely used, for example, in the presently expanding digital video and audio fields. In fact, digital video and audio applications require higher and higher storage capacities in order to store a large number of musical songs in the same support, or to enhance image quality, such as by an increased number of imaging pixels.

[0008] Multi-level non-volatile memories have recently appeared on the market, which are memories capable of storing a multiplicity of information bits in each cell. Such memories look specially well equipped to fill the above demands.

[0009] In this type of multi-level memories, the charge stored in the floating gate is further broken up, thereby generating a number 2^{nb} of distributions, where "nb" is the number of bits that are to be stored in a single cell. For example, with 2 bits per cell, the read sense amplifier is to process four distributions, instead of two as in the two-level instance.

[0010] A comparison of the threshold voltage distribution for a two-level memory and a multi-level memory with two bits per cell is schematically illustrated in Figures 3A and 3B. It can be seen that the multi-level structure decreases the gap between the voltage values and increases the read voltage.

[0011] It should be noted that the working range of the threshold voltage is independent of the number of bits contained in the cell. Thus, by employing a multi-level structure, the threshold gap between the various distributions decreases.

[0012] Reducing the gap between the threshold voltage distributions means reducing the current differences that the sense amplifier is to sense. Furthermore, it should be provided a programming method, able to place the cells inside the various distributions.

[0013] For convenience of illustration only, the instance of a flash EEPROM with NOR architecture will be considered herein below.

[0014] As it is well known, memory cells of this type are written by hot electron injection, by applying a voltage of about 10 V to the control gate terminal , a voltage of about 5 V to the drain terminal , and by leaving the source terminal connected to a ground reference, thereby allowing the floating gate terminal to accumulate charge to saturation.

[0015] In the instance of a multi-level memory, the reduction of the difference between the threshold voltages corresponding to the various charge levels that can be stored in the floating gate terminal, and hence the difference between the various conduction levels of the cells, requires an accurate and "fine" control of the cell programming operation, and in particular of the charge stored in the floating gate terminal during such an operation.

[0016] It has been shown, both theoretically and experimentally, that a linear relation exists between the variation ΔVG of the voltage applied to the control gate terminal during the cell programming phase and the threshold jump that is obtained at set values of the voltage VD applied to the drain terminal and of the voltage VS to the source terminal, as explained by Riccò et al. in an article "Nonvolatile multilevel memories for digital application", Pro. IEEE, vol. 86, December, 1998, pages 2399-2421.

[0017] In particular, as schematically shown in Figure

1, the cell should be programmed by applying, to its control gate terminal, a "stepwise" voltage that increases linearly.

[0018] In practice, a series of program pulses are used which differ from each other by a constant value ΔVG. Thus, the program voltage is a constant pitch stepwise ramp, while the voltage on the drain terminal and the pulse duration are dependent on and set by the cell fabrication process.

[0019] At the end of each program pulse, the result is verified to see if the desired threshold level has been attained, and to discontinue or to continue programming accordingly.

[0020] It can be appreciated that, when this programming method is used, a threshold voltage distribution of width ΔVG is obtained, which equals the pitch of the stepwise program voltage.

[0021] Thus, multi-level memory cells can be programmed at a desired threshold voltage by using a predetermined number of program pulses.

[0022] The main problem encountered with the above method is its inherently low speed. Programming multi-level cells involves applying a succession of pulses to the control gate of the cell, starting from the lowest level, and this takes a longer time than the single programming pulse used in the case of two-level cells. In addition, each level is attained only after the setting of the level directly below.

[0023] In order to achieve a programming time of the single byte that can be compared with that of a conventional two-level cell, it has been thought of programming several multi-level cells in parallel.

[0024] Assuming 8 μs to be the time taken to program a single byte in the two-level case, and 200 μs to be the time taken to go through the stepwise program ramp in the multi-level case, then 256 bits of multi-level cells would have to be programmed simultaneously in order to achieve an effective programming time of 6 μs per each single byte of multi-level cells.

[0025] Increasing the internal parallelism of multi-level memory devices would bring about several technical problems, first and foremost increased current usage, to the point that any engineering developments in that direction would be restrained.

[0026] The underlying technical problem of this invention is to provide a new method of programming the cells of a multi-level non-volatile memory device, which method has appropriate operational features to enable programming to be performed faster than, yet as accurately as, the programming methods of the prior art.

Summary of the Invention

[0027] The solution idea of the present invention is that of applying, to the drain terminals of the cells of a given memory word to be programmed, different voltage values as a function of of the threshold to be attained. The different drain voltages, each corresponding to a predetermined level, are selected to promote the attainment of the corresponding level in a seeking-to or asymptotic way, i.e. after a congruous number of pulses, regardless of the final level of each bit.

[0028] Of course, to obtain bits of equal level, the same drain voltage values have to be applied.

[0029] The congruous number of pulses is to meet two requirements: it should be the least possible, and at the same time, ensure that each level is attained with an appropriate degree of suitable and controlled accuracy.

[0030] Based on this idea, the technical problem is solved by a programming method as previously indicated and as defined in Claim 1.

[0031] The features and advantages of the programming method according to this invention will become apparent from the following description of an embodiment thereof, given by way of non-limitative example with reference to the accompanying drawings.

Brief Description of the Drawings

[0032] In the drawings:

- Figure 1 schematically shows schematically a series of multi-level memory cells and a plot of a programming voltage to be applied to the cell control gate terminals, according to the prior art;
- Figure 2 schematically shows the same series of multi-level memory cells as in Figure 1, and a plot of a programming voltage to be applied to the cell drain terminals, according to this invention;
- Figure 3 schematically shows a plot of voltage against time showing certain characteristic curves that correspond to different threshold levels of the memory cells;
- Figures 3A and 3B are comparative of the threshold voltage distribution, in a two-level memory and a multi-level memory having two bits per cell;
- Figures 4A and 4B schematically show voltage vs. time plots respectively showing two different characteristic curves associated with a predetermined logic level that vary depending on the number of pulses applied to the drain terminals of the memory cells; and
- Figures 5A and 5B schematically show voltage vs. time plots respectively showing two different characteristic curves associated with a predetermined logic level and that vary depending on the number of pulses applied to the drain terminals of the memory cells.

Detailed Description

[0033] With reference to the drawings, in particular to Figure 9 thereof, a group of memory cells, being incorporated to a multi-level non-volatile memory device monolithically integrated in a semiconductor, are shown generally at 1 in schematic form.

[0034] The memory device is a conventional one, and accordingly shown in the drawings as a whole. It comprises an array of multi-level memory cells along with its address, decode, read, write, and erase circuitry. Thus, shown in Figure 1 is just an eight-bit memory word or byte.

[0035] However, there is no reason why larger memory words, e.g. words of sixteen, thirty-two, or sixty-four bits, could not be considered instead. The following considerations bear no relation to the size of the memory word.

[0036] The memory device where the cells 1 are incorporated may be of the EEPROM or of the flash EEPROM, and is electrically programmable and erasable.

[0037] As previously stated, the array of non-volatile memory cells has conventional decode circuitry associated therewith, which circuitry is triggered by a switching edge of a pulsive signal ATD.

[0038] A typical architecture of a non-volatile memory array comprises a number of sectors divided into two groups. Each sector can be programmed and erased independently of the other sectors. Furthermore, each group has its own row decode, and its own array of read sense amplifiers.

[0039] A more detailed description of the construction and operation of a multi-level memory device can be found in the technical literature. For the purpose of this invention, it is important to understand that, in a multi-level memory device, the duration of the programming pulse and the voltage applied to the drain terminal of the memory cell are critical parameters to successful programming.

[0040] In conventional programming of the stepwise type, carried out according to the prior art, a memory cell typically draws a current of several tens of μA. For example, assuming a current draw of 50 μA per cell, programming or modifying 64 cells in parallel would require a total program current of 3.2 mA.

[0041] The method of the invention allows the current requirements for programming to be drastically reduced, and allows much faster programming.

[0042] With reference in particular to the example shown in Figure 2, the control gate terminals of the memory cells 1 that comprise the memory word are all held at the same constant voltage value Vg.

[0043] At the same time, to the drain terminal of each memory cell is applied a drain voltage that is a direct function of the threshold level sought to be reached for a predetermined bit. The voltage values to be applied to the drain terminal, say four different values for four different levels, are purposely selected from an indefinite set of possible values, such that the target condition can be simultaneously attained, regardless of the values of the levels. This is achieved with a limited, albeit exactly congruous, number of pulses, which are followed by a testing phase.

[0044] This is made practicable by the drain levels being designed to produce larger variations at higher levels, such that they can attain a respective level each at the end of a convenient number of pulses.

[0045] In this way, a different final threshold voltage can be obtained for each cell at the very instant of time when programming is discontinued.

[0046] The aspects of the invention can be better appreciated by having reference to the graph of voltage vs. time in Figure 3.

[0047] The graph shows curves of measurements made by the Applicant on electrically programmable/erasable multi-level non-volatile cells.

[0048] The curves show that the write characteristic of the cells varies with time and the value of the drain voltage at constant gate voltage, e.g. of 9 V.

[0049] The cell threshold increases with time and according to the end level sought, while at a predetermined instant of time, it takes a value that corresponds to the drain voltage applied.

[0050] In essence, after a predetermined period of time, the threshold voltage of a given cell held at a constant control gate voltage will have a threshold value that is tightly linked to the voltage applied to its drain terminal. The higher this voltage value, the higher will be the corresponding threshold. The drain voltage values are selected such that the different levels will be attained simultaneously after n congruously accurate pulses, whose number may be a bare minimum.

[0051] Of course, the minimum number of preceding pulses is tightly linked to the degree of accuracy sought. However, by paralleling the levels in this manner, some more pulses can be expended to enhance the degree of accuracy, with the inventive method providing the dual advantage of being faster and more accurate than traditional approaches.

[0052] Consequently, the method of this invention essentially dissociates the memory cell programming time from a target threshold level.

[0053] In conventional stepwise programming methods, programming time is a linear function of the level to which a given cell is to be raised, and many pulses are expended to just move from one level to another, with an attendant waste of time.

[0054] In addition, as the plot of Figure 3 brings out, attaining a given threshold level in a non-linear, or rather logarithmic fashion versus time is assurance of the threshold levels, and hence the corresponding logic levels, being neatly separated. In essence, the possibility for levels to overlap is drastically lessened.

[0055] A noteworthy feature of the programming method of this invention is that to the drain terminals of the cells are applied successive voltage pulses of pre-

determined duration, e.g. 1 μs, and of a level same as the level sought. However, there is no reason why successive pulses of different duration could not be used instead, e.g. a first pulse of long duration followed by one or more pulses of shorter or decreasing duration, in order to advantageously speed up the starting step and further reduce the number of pulses and attendant verify step.

[0056] The voltage/time graph of Figure 4A shows, by way of example, two characteristic curves illustrating how the attainable threshold values may vary for the same applied voltage values but with a different number of pulses.

[0057] The voltage/time graph of Figure 4B is another example obtained at a different drain voltage level if compared with the graph of Figure 4A.

[0058] The graphs of Figures 5A and 5B are similar to those of Figures 4A and 4B, but for the number of pulses that are applied in order to attain a predetermined threshold.

[0059] The number of pulses that can be applied to the drain terminal may vary contingent on the cell fabrication process, and can be adjusted according to the results of conventional program verify tests.

[0060] The method of this invention does brilliantly solve the technical problem, and provides a number of advantages, foremost among which is that the number of pulses, and with it the memory word programming time, has been greatly reduced if compared with conventional methods.

[0061] Simulation tests conducted by the Applicant have demonstrated time savings of up to one order of magnitude over gate pulse writing.

[0062] The solution proposed by the present invention outdates the parallelism techniques.

[0063] An additional advantage comes from the reduced stress on the gates of the memory cells. In the state of the art, one is to wait for all the cells of a byte to be written before a memory address can be changed, which highly stresses the gate terminals with the application of the whole stepwise ramp of voltage values. On the other hand, the amount of stress on the gate terminals is reduced in this invention, again of one order of magnitude.

[0064] Finally, it will be appreciated that, in this invention, the drain terminals of the cells to be programmed are brought up to a desired regulated voltage value, regardless of the number of cells that are to be programmed.

[0065] The programming method of this invention can also be advantageously applied to high-capacity memory devices having special functions (such as the burst mode or page mode) and including a large number of sense amplifiers arranged for sharing by the whole memory array.

Claims

1. A method for programming a non-volatile memory device of the multi-level type, comprising a plurality of transistor cells grouped into memory words and provided with conventional gate and drain terminals, **characterized in that** different drain voltage values are applied in parallel to separate cells for attaining different threshold values.

2. A programming method according to Claim 1, **characterized in that** said different drain voltage values are directly proportional to the threshold values to be at tained for each bit, and are selected such that the corresponding level set for each bit is attained in a seeking-to or asymptotic manner at the end of a congruous number of pulses.

3. A programming method according to Claim 1, **characterized in that** a constant gate voltage value is concurrently applied to the gate terminals of said cells.

4. A programming method according to Claim 1, **characterized in that** each different drain voltage value is applied by means of a succession of pulses.

5. A programming method according to Claim 4, **characterized in that** the different drain voltages have a constant amplitude or level at each pulse.

6. A programming method according to Claim 4, **characterized in that** said pulses are constant duration pulses.

7. A programming method according to Claim 4, **characterized in that** said pulses are decreasing duration pulses.

8. A programming method according to Claim 4, **characterized in that** said pulses are varying duration pulses.

9. A programming method according to Claim 1, **characterized in that** the different drain voltage values correspond to different logic threshold values of the multi-level cells.

10. A programming method according to Claim 2, **characterized in that** said different drain voltage values are appropriate to ensure that a predetermined threshold level will be attained in a seeking-to manner, at the end of a predetermined number of pulses.

11. A programming method according to Claim 1, **characterized in that** said some cells of said memory word take a different individual threshold value in a

seeking-to manner, at the same instant of time when programming is interrupted.

12. A programming method according to Claim 1, **characterized in that** the cell programming time is unrelated to the threshold level sought.

13. A programming method according to Claim 2, **characterized in that** the point of simultaneous attainment in a seeking-to manner of the different target levels is set at the end of a limited number of pulses.

14. A programming method according to Claim 13, **characterized in that** the limited number of pulses is set consistently with a congruous degree of accuracy and with a number tending to a minimum.

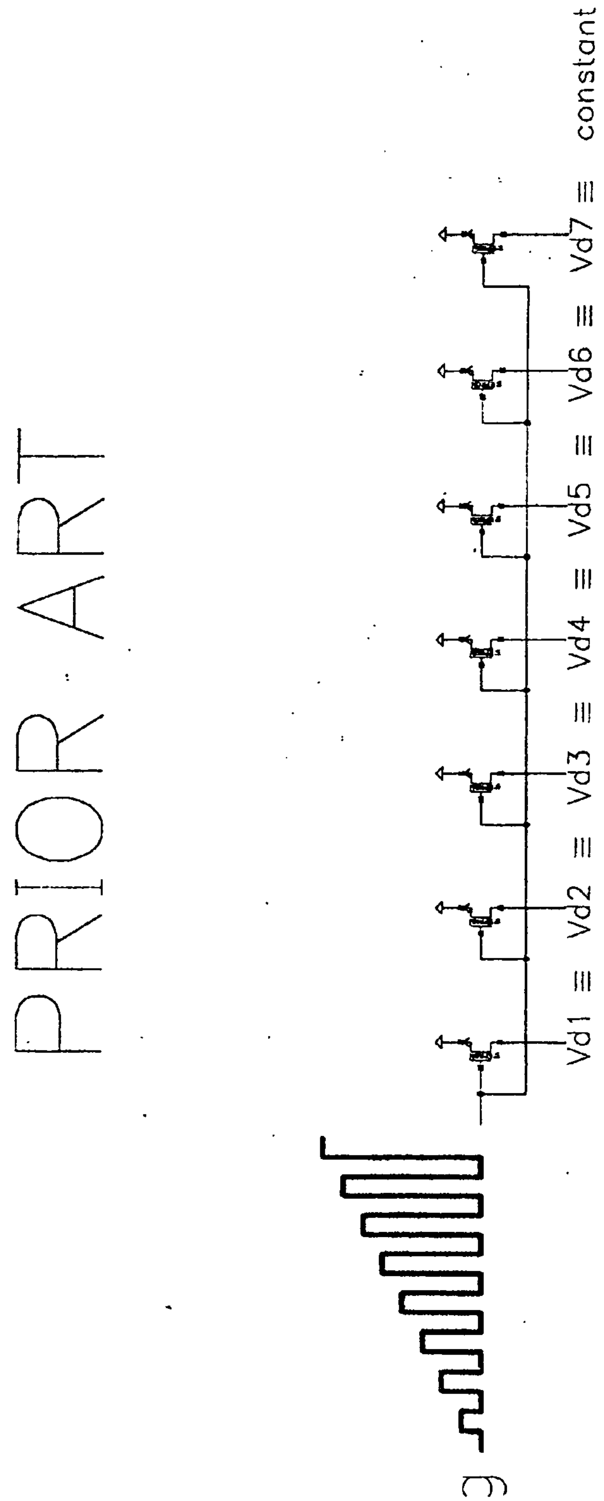


FIG. 1

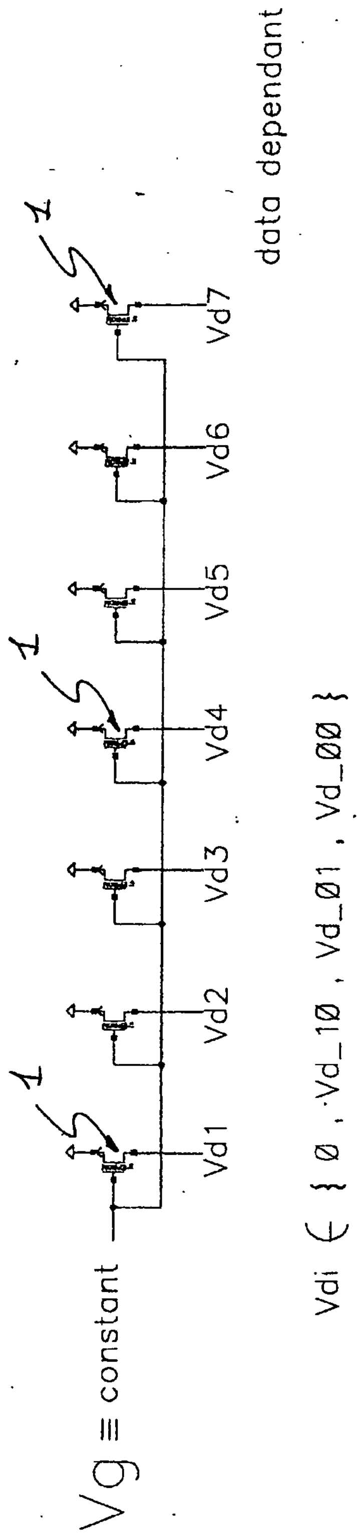

Vg ≡ constant
1
1
1
Vd1
Vd2
Vd3
Vd4
Vd5
Vd6
Vd7
data dependant
Vdi ∈ { Ø , Vd_1Ø , Vd_Ø1 , Vd_ØØ }

FIG. 2

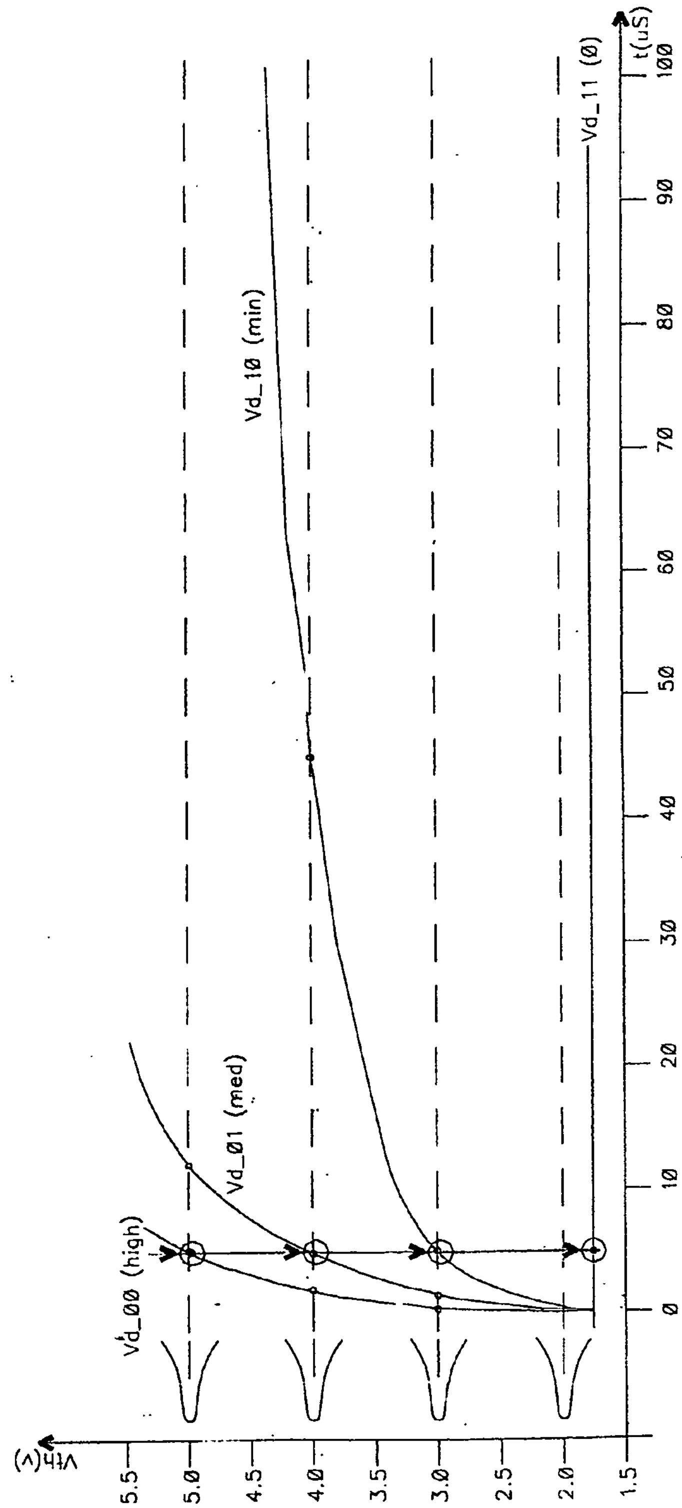
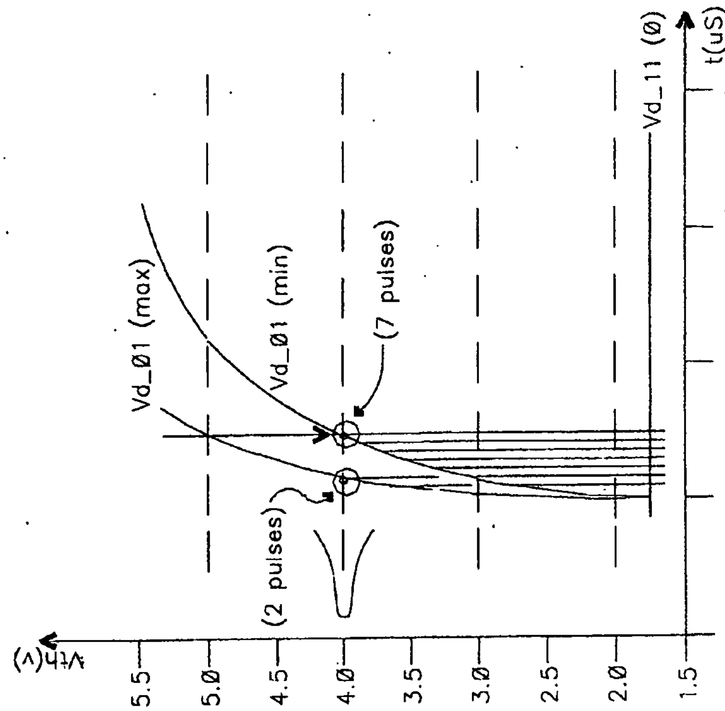
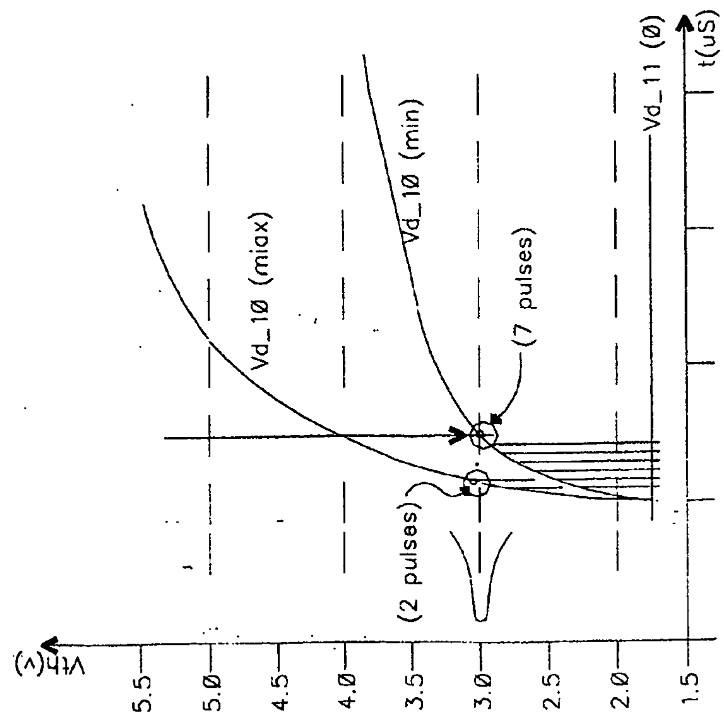

Vth(V)
5.5
5.Ø
4.5
4.Ø
3.5
3.Ø
2.5
2.Ø
1.5
t(uS)
Ø
1Ø
2Ø
3Ø
4Ø
5Ø
6Ø
7Ø
8Ø
9Ø
1ØØ
Vd_ØØ (high)
Vd_Ø1 (med)
Vd_1Ø (min)
Vd_11 (Ø)

FIG. 3



Vth(v)
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
Vd_Ø1 (max)
Vd_Ø1 (min)
(7 pulses)
(2 pulses)
Vd_11 (Ø)
t(uS)
Ø
1Ø
2Ø
3Ø

FIG. 4B



Vth(v)
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
Vd_1Ø (miax)
Vd_1Ø (min)
(7 pulses)
(2 pulses)
Vd_11 (Ø)
t(uS)
Ø
1Ø
2Ø
3Ø

FIG. 4A

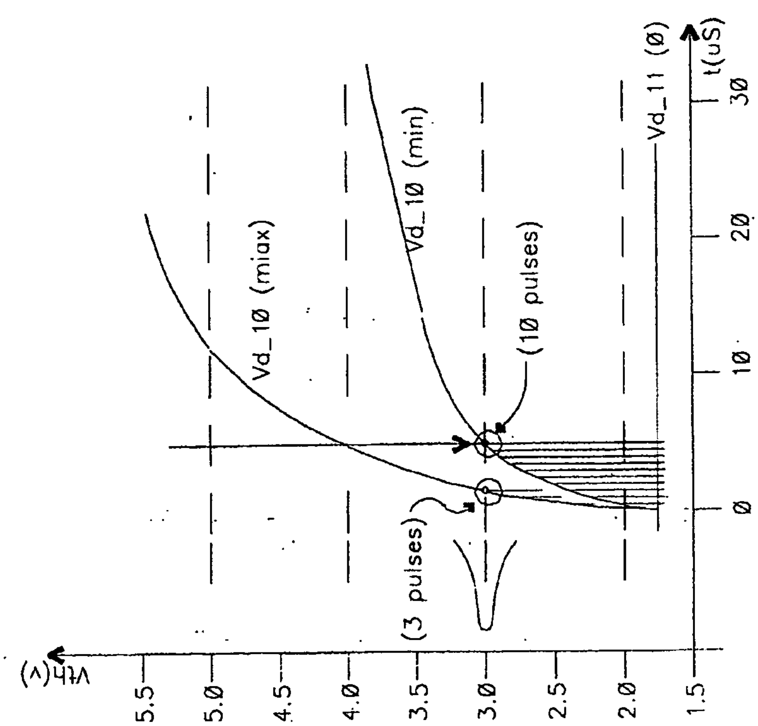

Vth(V)
5.5
5.Ø
4.5
4.Ø
3.5
3.Ø
2.5
2.Ø
1.5
t(uS)
Ø
1Ø
2Ø
3Ø
Vd_1Ø (miax)
Vd_1Ø (min)
(3 pulses)
(1Ø pulses)
Vd_11 (Ø)

FIG. 5A

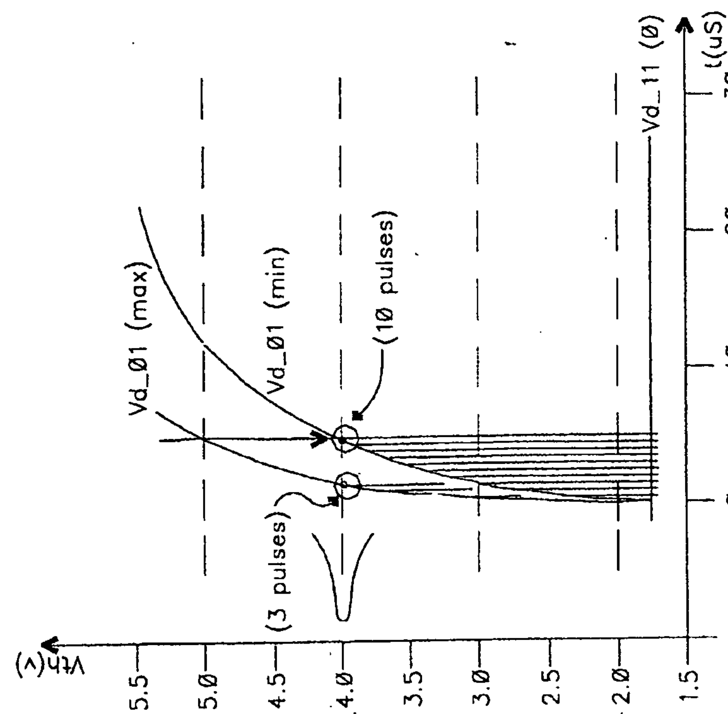

Vth(V)
5.5
5.Ø
4.5
4.Ø
3.5
3.Ø
2.5
2.Ø
1.5
t(uS)
Ø
1Ø
2Ø
3Ø
Vd_Ø1 (max)
Vd_Ø1 (min)
(3 pulses)
(1Ø pulses)
Vd_11 (Ø)

FIG. 5B

European Patent Office

EUROPEAN SEARCH REPORT

Application Number
EP 02 42 5293

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.7)
X Y	US 5 970 012 A (TAKESHIMA TOSHIO) 19 October 1999 (1999-10-19) * column 4, line 44 - column 8, line 9 * * figures 1-7 *	1,3,9, 11,12 2,4-8, 10,13,14	G11C16/10 G11C11/56
X	US 5 796 652 A (SUGAWARA HIROSHI ET AL) 18 August 1998 (1998-08-18) * column 6, line 28 - column 9, line 16 * * figures 8-10 *	1,3,9, 11,12	
X	US 5 708 600 A (SATO SHINICHI ET AL) 13 January 1998 (1998-01-13) * column 6, line 62 - column 9, line 48 * * figures 1,2,4 *	1,3,9,12	
Y	US 5 553 020 A (ATWOOD GREGORY E ET AL) 3 September 1996 (1996-09-03) * column 2, line 41 - column 3, line 18 * * column 7, line 34 - line 60 * * figure 10 *	2,4-8, 10,13,14	

TECHNICAL FIELDS SEARCHED (Int.Cl.7)

G11C

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
THE HAGUE	24 September 2002	Colling, P

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)

ANNEX TO THE EUROPEAN SEARCH REPORT ON EUROPEAN PATENT APPLICATION NO. EP 02 42 5293

This annex lists the patent family members relating to the patent documents cited in the above-mentioned European search report.
The members are as contained in the European Patent Office EDP file on
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

24-09-2002

Patent document cited in search report		Publication date	Patent family member(s)		Publication date
US 5970012	A	19-10-1999	JP	3204119 B2	04-09-2001
			JP	10106277 A	24-04-1998
US 5796652	A	18-08-1998	JP	9180473 A	11-07-1997
US 5708600	A	13-01-1998	JP	9213085 A	15-08-1997
US 5553020	A	03-09-1996	US	5487033 A	23-01-1996

EPO FORM P0459

For more details about this annex : see Official Journal of the European Patent Office, No. 12/82